

MODEL REGRESI NONPARAMETRIK DENGAN PENDEKATAN DERET FOURIER PADA KASUS TINGKAT PENGANGGURAN TERBUKA DI JAWA TIMUR

Alan Prahutama

Dosen Jurusan Statistika Undip

alan.prahutama@gmail.com

Abstrak

Pendekatan regresi dapat dilakukan dengan tiga pendekatan yaitu pendekatan parametrik, semiparametrik, dan nonparametrik. Salah satu pendekatan nonparametrik yang dikembangkan adalah menggunakan deret Fourier. Regresi nonparametrik deret Fourier menghasilkan kurva sinus cosinus, sehingga sebaran data yang berulang sangat sesuai didekati menggunakan deret Fourier. Estimasi deret Fourier dapat menggunakan OLS (*Ordinary Least Square*). Pada estimasi deret Fourier penentuan K optimal menggunakan GCV (*Generalized Cross Validation*). Pengaplikasian pendekatan regresi nonparametrik menggunakan deret Fourier pada Tingkat Pengangguran terbuka di Jawa Timur. Variabel independennya adalah Presentase penduduk kerja usia 15 tahun keatas, angka partisipasi kasar, laju pertumbuhan ekonomi daerah, laju pertumbuhan penduduk daerah, presentase perusahaan, tingkat investasi, dan tingkat minimum. Hasil yang didapat bahwa model menghasilkan R-square 96.76% dan merupakan parsimoni model.

Kata Kunci: Regresi Nonparametrik, Deret Fourier, GCV, Pengangguran Terbuka di Jawa Timur.

1. Pendahuluan

Metode regresi merupakan metode statistika untuk mengetahui hubungan antara variabel dependen dengan variabel independen. Pendekatan regresi dapat dilakukan dengan tiga pendekatan yaitu parametrik, semiparametrik dan nonparametrik. Suatu pendekatan nonparametrik dilakukan apabila fungsi dari regresi tersebut tidak diketahui bentuk kurva fungsinya (Budiantara, 2000). Pendekatan regresi nonparametrik telah banyak dikembangkan antara lain menggunakan Spline, Kernel, Polinomial Lokal, Wavelet, dan Fourier. Salah satu keunggulan pendekatan regresi nonparametrik dengan menggunakan deret Fourier adalah mampu mengatasi data yang mempunyai sebaran trigonometri, dalam hal ini adalah sinus dan cosinus. Pola data yang sesuai dengan pendekatan Fourier merupakan pola data yang berulang, yaitu pengulangan terhadap nilai variabel dependen untuk variabel independen yang berbeda-beda. Penelitian menggunakan deret Fourier untuk nonparametrik maupun semiparametrik telah banyak dikembangkan. Penelitian mengenai deret Fourier antara lain Bilodeau (1992), kemudian Tripena (2007) mengkaji estimator deret Fourier pada regresi nonparametrik, Semiati (2010) mengembangkan estimasi model regresi nonparametrik deret Fourier biresponse. Sedangkan untuk regresi semiparametrik menggunakan deret Fourier telah dikembangkan oleh Asrini (2011).

Indonesia merupakan Negara berkembang dimana salah satu permasalahan yang dihadapi adalah tingkat pengangguran. Tingkat pengangguran di Indonesia cukup

tinggi, hal ini dikarenakan minimnya lapangan pekerjaan sedangkan para pencari kerja meningkat. Tingkat pengangguran di Indonesia pada tahun 2010 mencapai 7,14%. Jawa Timur merupakan salah satu provinsi di Indonesia yang padat penduduknya. Jumlah penduduk di Jawa Timur tahun 2010 mencapai 37.476.491 jiwa (BPS, 2010). Jumlah pengangguran di Jawa Timur pada tahun 2011 mencapai 821.546 jiwa hal tersebut didasarkan bahwa angkatan kerja pada tahun 2011 mencapai 19.761 juta jiwa sedangkan jumlah tenaga kerja yang terserap sebesar 18.940 juta jiwa. Penelitian mengenai pengangguran terbuka telah banyak dilakukan antara lain Asti (2011) melakukan pengelompokan mengenai faktor-faktor yang mempengaruhi tingkat pengangguran terbuka di Jawa Timur dengan metode MARS. Sari dan Budiantara (2012) memodelkan pengangguran terbuka di Jawa Timur dengan pendekatan Spline. Oleh karena itu peneliti ingin meneliti mengenai pemodelan pengangguran terbuka di Jawa Timur menggunakan deret Fourier. Hal ini didasarkan pada pola sebaran data tingkat pengangguran terbuka di Jawa Timur terhadap faktor-faktor yang mempengaruhinya menyebar secara berulang.

2. Landasan Teori

Landasan teori yang digunakan adalah analisis regresi, regresi nonparametrik, deret Fourier, GCV (*Generalized Cross Validation*) dan Pengangguran Terbuka.

2.1 Analisis Regresi

Analisis regresi adalah analisis hubungan antara variabel dependen dengan variabel independen. Model persamaan regresi adalah sebagai berikut:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \varepsilon_i \quad (1)$$

dengan y_i merupakan variabel dependen sedangkan x merupakan variabel independen. Pada regresi parametrik residual diasumsikan identik, independen, dan berdistribusi Normal $\varepsilon \sim IIDN(0, \sigma^2)$. Estimasi parameter model regresi menggunakan metode OLS (*Ordinary Least Square*) atau menggunakan metode MLE (Drapper dan Smith, 1992).

2.2 Regresi Nonparametrik

Regresi nonparametrik merupakan pendekatan metode regresi dimana bentuk kurva dari fungsi regresinya tidak diketahui. Kurva fungsi diasumsikan termuat dalam ruang fungsi tertentu (Eubank, 1988). Model regresi nonparametrik adalah sebagai berikut:

$$y_i = \eta(x_i) + \varepsilon_i \quad (2)$$

$\eta(x_i)$ merupakan kurva fungsi regresi yang tidak diketahui bentuknya dengan x_i merupakan variabel independen. Kurva fungsi $\eta(x_i)$ diasumsikan *smooth* pada ruang fungsi tertentu.

2.3 Deret Fourier

Deret Fourier adalah fungsi polinomial trigonometri yang mempunyai tingkat fleksibilitas. Hal ini dikarenakan bahwa deret Fourier merupakan kurva yang menunjukkan fungsi sinus dan cosinus. Fungsi dari deret Fourier adalah sebagai berikut:

$$f(x) = \frac{1}{2} \alpha_0 + \gamma x + \sum_{k=1}^K \alpha_k \cos kx \quad (3)$$

(Bilodeau, 1992).

2.4 GCV (Generalized Cross Validation)

Pada pemodelan regresi nonparametrik dengan menggunakan deret Fourier, hal yang perlu diperhatikan adalah menentukan nilai K. Penentuan K optimal bisa menggunakan metode GCV (*Generalized Cross Validation*). Penentuan K optimal akan menghasilkan nilai R^2 yang tinggi. Menurut Wu dan Zhang (1988) nilai GCV didapat sebagai berikut:

$$GCV(K) = \frac{MSE(K)}{(n^{-1} \text{trace}(\mathbf{I} - \mathbf{A}))^2} \quad (4)$$

$$\text{dengan } MSE(K) = n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (5)$$

Nilai GCV terkecil akan menghasilkan nilai K yang optimal.

2.5 Pengangguran Terbuka

Pengangguran terbuka merupakan bagian dari angkatan kerja yang tidak bekerja atau sedang mencari pekerjaan atau sedang mempersiapkan suatu usaha (BPS, 2013). Angka partisipasi kasar adalah rasio jumlah siswa, berapapun usianya yang sedang sekolah di tingkat pendidikan tertentu terhadap jumlah penduduk kelompok usia yang berkaitan dengan jenjang pendidikan tertentu. Kegunaan dari APK adalah mengukur daya serap usia sekolah di setiap jenjang pendidikan. Laju pertumbuhan ekonomi merupakan laju pertumbuhan yang dibentuk dari berbagai sektor ekonomi yang secara tidak langsung menggambarkan tingkat pertumbuhan ekonomi yang terjadi di daerah tersebut. Kegunaan mengetahui laju pertumbuhan ekonomi adalah mengetahui keberhasilan tingkat ekonomi di masa mendatang (BPS, 2013). Jika laju pertumbuhan ekonomi semakin tinggi, mengindikasikan bahwa tingkat pengangguran semakin rendah. Presentase perusahaan dan tingkat investasi di suatu kabupaten/kota menunjukkan tingkat pengangguran di kabupaten/kota tersebut.

3. Metodologi Penelitian

Langkah-langkah yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Menentukan estimator regresi nonparametrik dengan pendekatan deret Fourier menggunakan metode OLS
2. Membuat program penentuan K optimal dengan metode GCV

3. Setelah didapat K optimal, selanjutnya adalah membuat program estimasi model regresi nonparametrik dengan pendekatan deret Fourier

Penelitian ini diaplikasikan pada tingkat pengangguran terbuka di Jawa Timur. Data yang diambil adalah data sekunder dari Sakernas (Satuan Kerja Nasional) 2010 BPS Jawa Timur Variabel-variabel yang digunakan dalam penelitian adalah sebagai berikut

Y = Tingkat pengangguran terbuka di Jawa Timur

X_1 = Presentase penduduk usia kerja berumur 15 tahun keatas berdasarkan pendidikan tertinggi yang ditamatkan

X_2 = Angka partisipasi Kasar menurut kabupaten/kota

X_3 = Laju pertumbuhan ekonomi daerah

X_4 = Laju pertumbuhan penduduk menurut kabupaten/kota

X_5 = Presentase perusahaan menurut kabupaten/kota

X_6 = Tingkat investasi menurut kabupaten/kota

X_7 = Tingkat upah minimum menurut kabupaten/kota

Software yang digunakan dalam penelitian ini adalah R.

4. Pembahasan

Pembahasan pada penelitian ini meliputi estimasi parameter model regresi nonparametrik dengan menggunakan deret Fourier dan pemodelan pengangguran terbuka di Jawa Timur menggunakan deret Fourier.

4.1 Estimasi parameter regresi nonparametrik deret Fourier

Diberikan data berpasangan sebagai berikut (y_j, x_{ij}) dengan $j = 1, 2, \dots, m$ menyatakan banyaknya pengamatan dan $i = 1, 2, \dots, n$ menyatakan banyaknya variabel independen. Model regresi adalah sebagai berikut:

$$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \dots + \beta_n x_{nj} + \varepsilon_j. \quad (6)$$

Jika Persamaan 6 dibuat dalam bentuk matriks maka

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

dengan

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}_{m \times 1}; \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{n1} \\ 1 & x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1m} & x_{2m} & \cdots & x_{nm} \end{bmatrix}_{m \times (n+1)}; \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}_{(n+1) \times 1}; \boldsymbol{\varepsilon} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix}_{m \times 1}$$

jika $\mathbf{y} = \mathbf{f}(\mathbf{x}) + \boldsymbol{\varepsilon}$ maka

$$\mathbf{f}(\mathbf{x}) = [f(x_{11}) \ f(x_{12}) \ \cdots \ f(x_{1m}) \ f(x_{21}) \ f(x_{22}) \ \cdots \ f(x_{2m}) \ \cdots \ f(x_{n1}) \ f(x_{n2}) \ \cdots \ f(x_{nm})]^T \quad (7)$$

Dengan $f(x)$ merupakan kurva yang tidak diketahui bentuknya. $f(x)$ didekati dengan menggunakan deret Fourier sebagai berikut:

$$f(x) = \frac{1}{2}\alpha_0 + \gamma x + \sum_{k=1}^K \alpha_k \cos kx \quad (8)$$

sehingga elemen diagonal di Persamaan (7) didapat

$$f(x_{ij}) = \frac{1}{2} \alpha_0 + \gamma_j x_{ij} + (\alpha_{i1} \cos x_{ij} + \alpha_{i2} \cos 2x_{ij} + \dots + \alpha_{iK} \cos Kx_{ij}).$$

Jika $\mathbf{f}(\mathbf{x}) = \mathbf{A}\boldsymbol{\theta}$ maka

$$\mathbf{A} = \begin{bmatrix} 1 & x_{11} & \cos x_{11} & \cos 2x_{11} & \dots & \cos Kx_{11} & x_{21} & \cos x_{21} & \cos 2x_{21} & \dots & \cos Kx_{21} & \dots & x_{n1} & \cos x_{n1} & \dots & \cos Kx_{n1} \\ 1 & x_{12} & \cos x_{12} & \cos 2x_{12} & \dots & \cos Kx_{12} & x_{22} & \cos x_{22} & \cos 2x_{22} & \dots & \cos Kx_{22} & \dots & x_{n2} & \cos x_{n2} & \dots & \cos Kx_{n2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1m} & \cos x_{1m} & \cos 2x_{1m} & \dots & \cos Kx_{1m} & x_{2m} & \cos x_{2m} & \cos 2x_{2m} & \dots & \cos Kx_{2m} & \dots & x_{nm} & \cos x_{nm} & \dots & \cos Kx_{nm} \end{bmatrix}_{(n(k+1)+1) \times m}$$

$$\boldsymbol{\theta} = [\phi \quad \gamma_1 \quad \alpha_{11} \quad \alpha_{12} \quad \dots \quad \alpha_{1K} \quad \gamma_2 \quad \alpha_{21} \quad \alpha_{22} \quad \dots \quad \alpha_{2K} \quad \dots \quad \gamma_n \quad \alpha_{n1} \quad \alpha_{n2} \quad \dots \quad \alpha_{nK}]^T_{1 \times (n(K+1)+1)}$$

dengan $\phi = \frac{n}{2} \alpha_0$

Estimasi model regresi nonparametrik adalah dengan menggunakan OLS (*Ordinary Least Square*) sebagai berikut:

$$\boldsymbol{\psi}(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}$$

$$\hat{\boldsymbol{\theta}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y} \quad (9)$$

Persamaan (9) merupakan estimator regresi nonparametrik menggunakan pendekatan deret Fourier.

4.2 Pemodelan Pengangguran Terbuka di Jawa Timur dengan Menggunakan Deret Fourier.

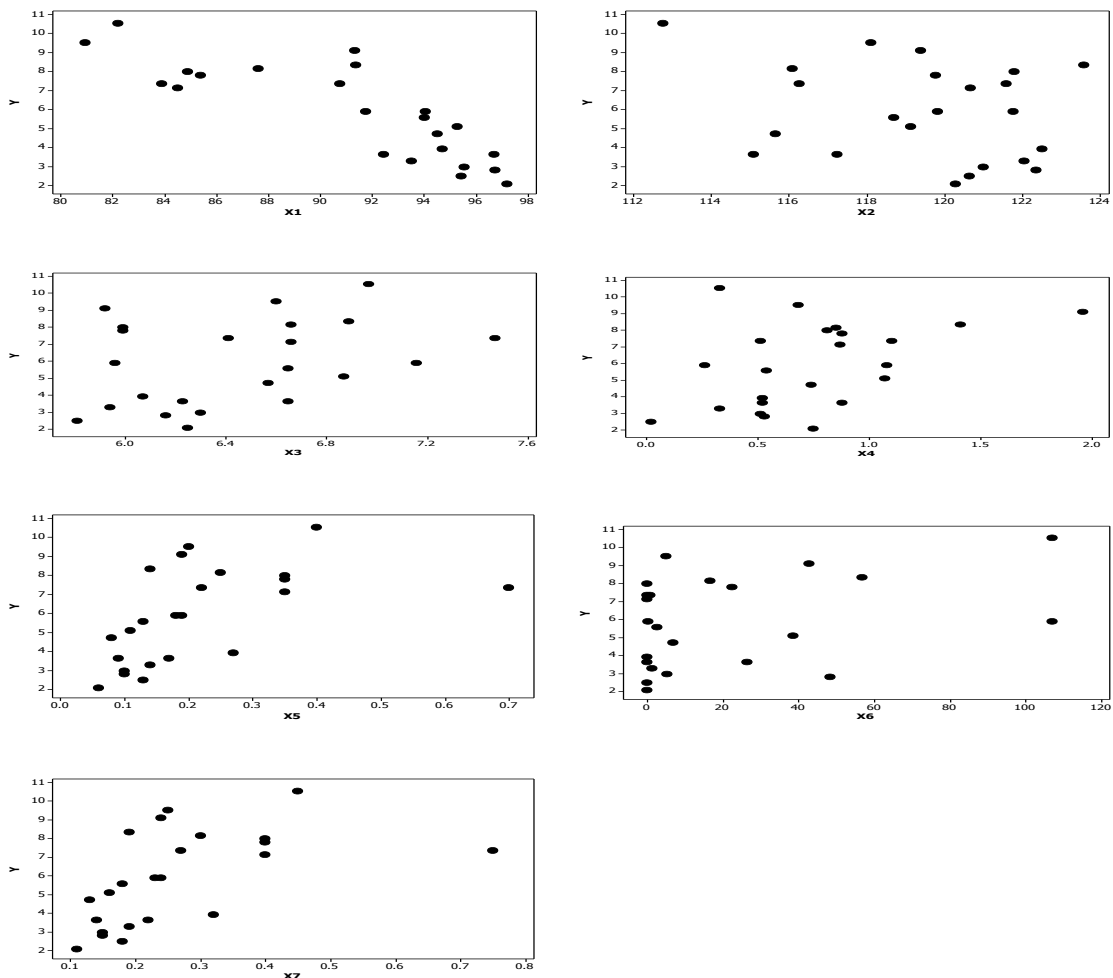
Gambar 1. merupakan sebaran variabel dependen yaitu tingkat pengangguran terbuka di Jawa Timur (y) terhadap presentase penduduk usia kerja 15 tahun keatas (x_1), angka partisipasi kasar (x_2), laju pertumbuhan ekonomi daerah (x_3), laju pertumbuhan penduduk menurut kabupaten/kota (x_4), presentase perusahaan menurut kabupaten/kota (x_5), tingkat investasi menurut kabupaten/kota (x_6), dan tingkat upah minimum menurut kabupaten/kota (x_7). Terlihat bahwa sebaran data acak dan berulang, sehingga digunakan pemodelan yang memiliki fleksibilitas.

Langkah pertama yang dilakukan adalah menentukan nilai K optimal. Nilai K merupakan bilangan bulat positif. Penentuan K optimal menggunakan metode GCV. Adapun hasil yang didapat dari setiap K yang dicobakan adalah sebagai berikut:

Tabel 1. Nilai GCV untuk setiap K

Nilai K	GCV	Nilai K	GCV	Nilai K	GCV
K=1	6.07×10^{-1}	K=6	1.05×10^{-23}	K=11	1.19×10^{-24}
K=2	6.04×10^{-2}	K=7	1.46×10^{-23}	K=12	6.82×10^{-25}
K=3	1.66×10^{-16}	K=8	7.3×10^{-24}	K=13	1.126×10^{-24}
K=4	1.57×10^{-21}	K=9	6.6×10^{-24}	K=14	9.6×10^{-25}
K=5	2.82×10^{-23}	K=10	2.22×10^{-24}	K=15	1.31×10^{-24}

Sumber: Olah Data dengan program R



Gambar 1. Scatter Plot antara Variabel Dependen dengan Variabel Independen
Berdasarkan Tabel 1 nilai GCV terkecil pada $K=12$. GCV terkecil merupakan K yang optimal. Jika nilai $K=12$ maka banyaknya parameter yang harus ditaksir sebanyak 92 parameter sehingga model yang dihasilkan tidak parsimoni. Oleh karena itu dilihat nilai R^2 nya untuk $K=1$ dan $K=2$.

Tabel 2.. Nilai R^2 dan MSE untuk setiap K

NILAI K	R^2	MSE
$K=1$	96.76%	0.194
$K=2$	99.3%	0.04
$K=12$	100%	6.24×10^{-25}

Sumber: Olah Data dengan program R

Berdasarkan Tabel 2 terlihat bahwa untuk nilai $K=1$ sudah menghasilkan nilai R^2 yang cukup tinggi yaitu sebesar 96.76%, sedangkan untuk nilai $K=2$ menghasilkan nilai R^2 sebesar 99.3%. Jika nilai K yang dipilih adalah $K=2$ maka estimasi parameter yang harus dicari adalah sebanyak 29 parameter. Model yang dipilih adalah model dengan R^2 yang besar dan juga model yang parsimoni (sederhana). Oleh karena itu model yang dipilih adalah $K=1$.

Tabel 3. Nilai Parameter untuk K=1

Parameter	Nilai	Parameter	Nilai	Parameter	Nilai
α_0	-62.4	γ_4	-0.229	α_{17}	4252.79
γ_1	-0.358	α_{14}	-3.84		
α_{11}	0.439	γ_5	116.45		
γ_2	-0.03	α_{15}	-4154.02		
α_{12}	-0.08	γ_6	0.018		
γ_3	0.742	α_{16}	-0.113		
α_{13}	1.66	γ_7	113.33		

Sumber: Olah Data dengan program R

Model yang didapat adalah sebagai berikut:

$$\hat{y} = -62.4 - 0.358x_1 + 0.439\cos x_1 - 0.03x_2 - 0.08\cos x_2 + 0.742x_3 + 1.66\cos x_3 - 0.229x_4 + \\ -3.84\cos x_4 + 116.45x_5 - 4154.02\cos x_5 + 0.018x_6 - 0.113\cos x_6 + 113.33x_7 + 4252.7\cos x_7$$

Nilai R^2 yang dihasilkan sebesar 96.76% .

5. Kesimpulan dan Saran

Estimasi parameter regresi dengan pendekatan deret Fourier menggunakan OLS menghasilkan $\hat{\theta} = (A^T A)^{-1} A^T y$. Pada pemodelan pengangguran terbuka di Jawa Timur ditentukan nilai K optimal dengan GCV. K optimal yang didapat yaitu K=12, sehingga jika digunakan K=12 menghasilkan model yang kompleks. Oleh karena itu dicobakan nilai K yang sederhana. Model K=1 menghasilkan model dengan $R^2 = 96.67\%$. Model yang dihasilkan adalah sebagai berikut:

$$\hat{y} = -62.4 - 0.358x_1 + 0.439\cos x_1 - 0.03x_2 - 0.08\cos x_2 + 0.742x_3 + 1.66\cos x_3 - 0.229x_4 + \\ -3.84\cos x_4 + 116.45x_5 - 4154.02\cos x_5 + 0.018x_6 - 0.113\cos x_6 + 113.33x_7 + 4252.7\cos x_7$$

Perkembangan penelitian mengenai deret Fourier dapat dikembangkan dengan meneliti sifat-sifat estimator dari deret Fourier. Perlu dikembangkan Estimasi multirespon regresi nonparametrik dengan pendekatan deret Fourier.

DAFTAR PUSTAKA

- Drapper, N.R dan Smith, H. 1992. *Applied Regression Analysis 2nd Edition*, Marcel Dekker, New York.
- Badan Pusat Statistika (BPS). 2013. "Ketenagakerjaan: Pengangguran Terbuka", www.bps.go.id di akses pada tanggal 12 Agustus 2013.
- Asrini, Luh Juni. 2012. "Regresi Parametrik Deret Fourier", *Prosiding Seminar Nasional FMIPA Universitas Negeri Surabaya*, hal.77-80, 24 November 2012.

- Budiantara, I.N. 2000. "Metode U, GML, CV, dan GCV dalam Regresi Nonparametrik Spline", *Majalah Ilmiah Himpunan Matematika Indonesia (MIHMI)*, Vol. 6, hal.285-290.
- Sari, R.S dan Budiantara, I.N. 2012. "Pemodelan Pengangguran Terbuka di Jawa Timur dengan Menggunakan Pendekatan Regresi Spline Multivariabel", *Jurnal Sains dan Seni ITS Vol. 1, No. 1*, hal. 236-241.
- Bilodeau, M. 1992. "Fourier Smoother and Additive Models", *The Canadian Journal of Statistics*, 3, hal. 257-259.
- Tripena, A. 2007. *Estimator Deret Fourier dalam Regresi Nonparametric*, Tesis, ITS, Surabaya.
- Semiati, Rini. 2010. *Regresi Nonparametrik Deret Fourier Birespon*, Tesis, ITS, Surabaya.
- Wu, H. dan Zhang, J.T. 2006. *Nonparametric Regression Methods for Longitudinal Data Analysis*, A John-Wiley and Son Inc. Publication, New Jersey.
- Eubank, R.L. 1988. *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, New York